

Identification and Estimation of Errors-in-Variables Using Nonnormality of the Unobserved Regressors

Dan Ben-Moshe*

The Hebrew University of Jerusalem

February 2014

Abstract

This paper identifies and estimates the coefficients in a multivariate errors-in-variables linear model when the unobserved arbitrarily dependent regressors are not jointly normal and independent of errors. To identify the coefficients, we use variation in the second-order partial derivatives of the log characteristic function of the unobserved regressors; a property of only not jointly normal distributions. A root- n consistent and asymptotically normal extremum estimator performs well in simulations relative to third and fourth order moment estimators.

Keywords: *Errors-in-variables, measurement errors, dependent regressors, nonnormality, second-order partial derivatives of the log characteristic function*

*Dan Ben-Moshe, Department of Economics, Hebrew University of Jerusalem, Jerusalem 91905, Israel. 972-2-5883365, danbm@huji.ac.il, <https://www.sites.google.com/site/dbmster>. I am grateful to Rosa Matzkin and Jinyong Hahn for their generous advice, support, and guidance. I thank Arthur Lewbel for helpful suggestions and references. The support of a grant from the Maurice Falk Institute for Economic Research in Israel is gratefully acknowledged.

1 Introduction

This paper investigates identification and estimation of a multivariate linear regression model with measurement error in all the regressors and arbitrarily dependent unobserved regressors using characteristic functions. We show that the coefficients are identified if the unobserved regressors are not jointly normal and are independent of the errors. Further, if any of the errors and regressors are dependent then the model is no longer identified.

The literature on errors-in-variables in linear models is vast. Three excellent reviews are Cheng and Van Ness (1999), Fuller (1987), and Gillard (2010). In the single regressor errors-in-variables model with independent regressor and errors, the coefficient is identified if and only if (a) the regressor is not normal or (b) neither of the errors are normal (Reiersøl, 1950; Schennach & Hu, 2013). In the multivariate errors-in-variables model with regressors that are independent of errors, it is known that the coefficients are identified if (a) the errors are normal and the regressors are not jointly normal (Willassen, 1979) or (b) the unobserved regressors have nonzero and finite third or higher order cumulants, which usually means that identification comes from skewness and / or kurtosis (Geary, 1949; Lewbel, 1997).

We show that the coefficients are identified when the unobserved regressors are not jointly normal, allowing the unobserved regressors to be arbitrarily dependent, the errors to have arbitrary distributions, and the third and higher order cumulants of all the random variables to be zero or infinite. Identification is based on the property that the second-order partial derivatives of the log characteristic function (LCF) of the unobserved regressors are not all equal to a constant if and only if they are not jointly normal. We will use this variation to show that the coefficients uniquely minimize a distance between second-order partial derivatives of LCFs and covariances of observables. Further, since the jointly normal distribution is the only distribution that does not have variation in all the second-order partial derivatives of its LCF, this provides a testable condition for

nonnormality. The identification strategy extends Ben-Moshe (2013) to allow dependent unobservables.

Estimation of the coefficients without any additional information is usually based on higher-order moments (Cragg, 1997; Dagenais & Dagenais, 1997; Erickson, Jiang, & Whited, 2013; Pal, 1980), which can have high variance and bias and are sensitive to outliers and data transformations. Our estimator is an extremum estimator based on second-order partial derivatives of the LCF of the observed variables, which contains all the information from the higher-order moments and is root- n consistent. In Monte-Carlo simulations, we find that our estimator performs well relative to estimators based on third and fourth order moments and is robust to various distributions including symmetric ones. Plots of the second-order partial derivatives of the LCF of the observables provide evidence for or against variation and appropriateness of using second-order partial derivatives for estimation, and evidence of not jointly normal unobserved regressors.

2 Theory

This section presents the model, assumptions, identification, and estimation. Our tool of analysis is the LCF, which is denoted by $\varphi_{\mathbf{X}^*}(\mathbf{s}) = \ln E[e^{i\sum_{k=1}^K X_k^* s_k}]$ where $\mathbf{X}^* = (X_1^*, \dots, X_K^*)$ is a random vector and $\mathbf{s} = (s_1, \dots, s_K) \in \mathbb{R}^K$.

2.1 The Model

Consider the classic errors-in-variables model,

$$\begin{aligned} X_k &= X_k^* + U_k & k = 1, \dots, K \\ Y &= \beta_1 X_1^* + \dots + \beta_K X_K^* + \varepsilon \end{aligned} \tag{1}$$

where Y is an observed outcome, $\mathbf{X} = (X_1, \dots, X_K)$ are observed measurements, $\mathbf{X}^* = (X_1^*, \dots, X_K^*)$ are unobserved regressors, $\mathbf{U} = (U_1, \dots, U_K)$ are unobserved measurement errors, ε is an unobserved error, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)$ are the unknown coefficients of interest. Instead of intercepts we allow $(\mathbf{X}^*, \mathbf{U}, \varepsilon)$ to have nonzero means. The identification strategy is invariant to these intercepts and means.

The following assumption describes the dependence structure, which allows the unobserved regressors to be arbitrarily dependent but independent of errors.

Assumption 2.1. *The unobserved errors U_1, \dots, U_K , and ε are mutually independent and independent of \mathbf{X}^* . The vector of unobserved regressors \mathbf{X}^* is dependent.*

No other unobserved variable can be dependent on \mathbf{X}^* for $\boldsymbol{\beta}$ to be identified.

Lemma 2.1. *If $(\mathbf{X}^*, \mathbf{U})$ or $(\mathbf{X}^*, \varepsilon)$ is dependent then $\boldsymbol{\beta}$ is not identified.*

For simplicity assume that \mathbf{X}^* cannot be divided into subsets of mutually independent vectors. If \mathbf{X}^* can be divided into mutually independent vectors then the following assumption is modified from \mathbf{X}^* is not jointly normal to each subset is not jointly normal. Even with the modification, the formula in Theorem 2.1 remains the same.

Assumption 2.2. *One of the following equivalent conditions hold:*

- (i) *The unobserved vector of regressors \mathbf{X}^* is not jointly normal,*
- (ii) *On all neighborhoods of the origin $\frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} \neq -\text{Cov}(X_{k_1}^*, X_{k_2}^*)$ for some (k_1, k_2) ,*
- (iii) *On all neighborhoods of the origin $\frac{\partial^2 \varphi_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_{k_2}} \Big|_{(\omega_1, \dots, \omega_K, 0)} \neq c_{k_1 k_2}$ for some (k_1, k_2) .*

In the single regressor errors-in-variables model, Reiersøl (1950) shows that a non-normal unobserved regressor is sufficient for identification and Schennach and Hu (2013) prove that the model is identified if and only if the unobserved regressor or both of the errors are nonnormal. In the multivariate errors-in-variables model, Willassen (1979) shows that if $(\mathbf{U}, \varepsilon)$ is jointly normal then $\boldsymbol{\beta}$ is identified if and only if \mathbf{X}^* is not jointly normal.

The intuition for requiring some nonnormality is that normal distributions are completely characterized by their means and covariances and Klepper and Leamer (1984) prove that higher-order moments (\mathbf{X}^*, Y) are necessary for identification.

While other papers identify β with restrictions on (and existence of) higher-order moments like nonzero skewness or kurtosis, we use all the information from higher-order moments and only require nonnormality of the unobserved regressors (Assumption 2.2(i)). Identification comes from variation in the second-order partial derivatives of the LCF of \mathbf{X}^* (Assumption 2.2(ii)), which is a characteristic of a random vector if and only if it is not jointly normal. Assumption 2.2(iii) is a testable restriction that can be checked in data and conceivably provides a test for normality.

2.2 Identification

Assuming that the unobserved regressors are not jointly normal and independent of the errors, the following theorem proves that the coefficients in the multivariate errors-in-variables model are identified.

Theorem 2.1. *Consider the multivariate errors-in-variables model from Equation (1). Let $\beta \in \mathbf{B} \subseteq \mathbb{R}^K$. If $\text{Var}(X_k^*) < \infty$ and $\beta_k \neq 0$ for $k = 1, \dots, K$, then β is identified when Assumptions 2.1 and 2.2 hold and is the unique solution to*

$$\beta = \underset{\mathbf{b} \in \mathbf{B}}{\text{argmin}} \int_{\mathbb{R}} \left(\sum_{1 \leq k_1 < k_2 \leq K} w_{k_1 k_2}(u) \left| \text{Cov}(X_{k_1}, X_{k_2}) + \frac{\partial^2 \varphi_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_{k_2}} \right|_{(b_1 u, \dots, b_K u, -u)} \right|^2 + \sum_{k_1=1}^K w_{k_1 Y}(u) \left| \text{Cov}(X_{k_1}, Y) + \frac{\partial^2 \varphi_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_Y} \right|_{(b_1 u, \dots, b_K u, -u)} \right|^2 \right) du \quad (2)$$

where $|z|^2 = |a + ib|^2 = a^2 + b^2$ is the absolute value of z and the weight functions $w_{k_1 \cdot}(u)$ satisfy $\int_{\mathbb{R}} w_{k_1 \cdot}(u) du = 1$ and $\int_{|u| < \delta} w_{k_1 \cdot}(u) du > 0$ for all $\delta > 0$.

Identification uses all cross partial derivatives and all the arguments of $\varphi_{\mathbf{X}, Y}$. It thus

seems hard to see how any assumption can be weakened and the coefficients still identified.

We sketch the proof with details in the appendix. The proof converts the problem to uniqueness of the solution of a functional equation. The LCF of the observed variables is,

$$\varphi_{\mathbf{X},Y}(\mathbf{s}) = \varphi_{\mathbf{X}^*}(s_1 + \beta_1 s_y, \dots, s_K + \beta_K s_y) + \varphi_\varepsilon(s_y) + \sum_{k=1}^K \varphi_{U_k}(s_k)$$

where the equality follows from the dependence structure in Assumption 2.2.

Identification comes from variation in the second-order partial derivatives of $\varphi_{\mathbf{X},Y}$ through choices of \mathbf{s} . Let Assumption 2.2(ii) hold for some fixed k_1 and k_2 with $k_1 \neq k_2$. Then,

$$\frac{\partial^2 \varphi_{\mathbf{X},Y}(\mathbf{s})}{\partial s_{k_1} \partial s_{k_2}} = \frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} \Big|_{(s_1 + \beta_1 s_y, \dots, s_K + \beta_K s_y)}$$

Substituting in $\mathbf{s} = \mathbf{0}$ and $\mathbf{s} = (b_1 u, \dots, b_K u, -u)$,

$$\text{Cov}(X_{k_1}, X_{k_2}) + \frac{\partial^2 \varphi_{\mathbf{X},Y}(\mathbf{s})}{\partial s_{k_1} \partial s_{k_2}} \Big|_{(b_1 u, \dots, b_K u, -u)} = \text{Cov}(X_{k_1}^*, X_{k_2}^*) + \frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} \Big|_{(u(b_1 - \beta_1), \dots, u(b_K - \beta_K))} \quad (3)$$

Only jointly normal random variables do not have variation in all second-order partial derivatives so that $\text{Cov}(X_{k_1}^*, X_{k_2}^*) + \frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} = 0$ for all $\boldsymbol{\omega}$ and all k_1 and k_2 . By Assumption 2.2(ii), Equation (3) equals zero if and only if $\{b_k = \beta_k\}_{k=1}^K$. Formula (2) follows by minimizing the distance between the partial derivatives of $\varphi_{\mathbf{X},Y}$ evaluated at $\mathbf{0}$ and $\mathbf{b} \in \mathbf{B}$, which by Assumption 2.2(ii) is uniquely minimized when $\mathbf{b} = \boldsymbol{\beta}$.

2.3 Estimation

For a given sample $\{Y_n, X_{1n}, \dots, X_{Kn}\}_{n=1}^N$ of iid observations, estimation replaces the population quantities in Equation (2) with sample analogs. The Extremum estimator is

$$\hat{\beta} = \underset{\mathbf{b} \in \mathbf{B}}{\operatorname{argmax}} \hat{Q}_N(\mathbf{b}) \quad (4)$$

where

$$\begin{aligned} \hat{Q}_N(\mathbf{b}) := & - \int_{\mathbb{R}} \left(\sum_{1 \leq k_1 < k_2 \leq K} \left| \widehat{\operatorname{Cov}}(X_{k_1}, X_{k_2}) + \frac{\partial^2 \widehat{\varphi}_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_{k_2}} \right|_{(b_1 u, \dots, b_K u, -u)} \right|^2 \\ & + \sum_{k_1=1}^K \left| \widehat{\operatorname{Cov}}(X_{k_1}, Y) + \frac{\partial^2 \widehat{\varphi}_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_y} \right|_{(b_1 u, \dots, b_K u, -u)} \right|^2 \Bigg) w(u) du \\ \\ \frac{\partial^2 \widehat{\varphi}_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_{k_2}} \Bigg|_{(b_1 u, \dots, b_K u, -u)} := & \frac{\left(\frac{1}{N} \sum_{n=1}^N X_{k_1 n} e^{iu(\sum_{k=1}^K b_k X_{kn} - Y)} \right) \left(\frac{1}{N} \sum_{n=1}^N X_{k_2 n} e^{iu(\sum_{k=1}^K b_k X_{kn} - Y)} \right)}{\left(\frac{1}{N} \sum_{n=1}^N e^{iu(\sum_{k=1}^K b_k X_{kn} - Y)} \right)^2} \\ & - \frac{\frac{1}{N} \sum_{n=1}^N X_{k_1 n} X_{k_2 n} e^{iu(\sum_{k=1}^K b_k X_{kn} - Y)}}{\frac{1}{N} \sum_{n=1}^N e^{iu(\sum_{k=1}^K b_k X_{kn} - Y)}} \\ \\ \frac{\partial^2 \widehat{\varphi}_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_y} \Bigg|_{(b_1 u, \dots, b_K u, -u)} := & \frac{\left(\frac{1}{N} \sum_{n=1}^N X_{k_1 n} e^{iu(\sum_{k=1}^K b_k X_{kn} - Y)} \right) \left(\frac{1}{N} \sum_{n=1}^N Y_n e^{iu(\sum_{k=1}^K b_k X_{kn} - Y)} \right)}{\left(\frac{1}{N} \sum_{n=1}^N e^{iu(\sum_{k=1}^K b_k X_{kn} - Y)} \right)^2} \\ & - \frac{\frac{1}{N} \sum_{n=1}^N X_{k_1 n} Y_n e^{iu(\sum_{k=1}^K b_k X_{kn} - Y)}}{\frac{1}{N} \sum_{n=1}^N e^{iu(\sum_{k=1}^K b_k X_{kn} - Y)}} \\ \\ \widehat{\operatorname{Cov}}(X_{k_1}, X_{k_2}) := & \frac{1}{N} \sum_{n=1}^N X_{k_1 n} X_{k_2 n} - \left(\frac{1}{N} \sum_{n=1}^N X_{k_1 n} \right) \left(\frac{1}{N} \sum_{n=1}^N X_{k_2 n} \right) \\ \\ \widehat{\operatorname{Cov}}(X_{k_1}, Y) := & \frac{1}{N} \sum_{n=1}^N X_{k_1 n} Y_n - \left(\frac{1}{N} \sum_{n=1}^N X_{k_1 n} \right) \left(\frac{1}{N} \sum_{n=1}^N Y_n \right) \end{aligned}$$

We repeat the following standard conditions and theorems needed for consistency and asymptotic normality of an extremum estimator (eg. Newey & McFadden, 1994).

Assumption 2.3. (i) $Q(\mathbf{b})$ is uniquely maximized at $\boldsymbol{\beta}$; (ii) $\boldsymbol{\beta} \in \text{int}(\mathbf{B})$ and \mathbf{B} is compact; (iii) $Q(\mathbf{b})$ is continuous; (iv) $\widehat{Q}_N(\mathbf{b})$ converges uniformly in probability to $Q(\mathbf{b})$; (v) $\widehat{Q}_N(\mathbf{b})$ is twice continuously differentiable in a neighborhood of $\boldsymbol{\beta}$; (vi) $\sqrt{N}\nabla_{\mathbf{b}}\widehat{Q}_N(\boldsymbol{\beta}) \xrightarrow{d} N(0, \Omega(\boldsymbol{\beta}))$; (vii) there is an $H(\mathbf{b})$ continuous at $\boldsymbol{\beta}$ such that $\nabla_{\mathbf{b}\mathbf{b}}\widehat{Q}_N(\mathbf{b})$ converges uniformly in probability to $H(\mathbf{b})$ in a neighborhood of $\boldsymbol{\beta}$; (viii) $H(\boldsymbol{\beta})$ is nonsingular.

Proposition 2.1. If Assumptions 2.3(i)-(iv) hold then $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$. If Assumptions 2.3(i)-(viii) hold then $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, H(\boldsymbol{\beta})^{-1}\Omega(\boldsymbol{\beta})H^{-1}(\boldsymbol{\beta}))$.

Assumption 2.3 is satisfied when $\boldsymbol{\beta}$ is identified i.e. Theorem 2.1 holds. Continuity and uniform convergence use the uniform continuity of a CF, assumptions that the CF is nonzero on an interval around the origin, and that moments are bounded. The proofs use Taylor expansions of sample analogs around population quantities. The exact expressions are lengthy but only require taking derivatives of the expressions above.¹

3 Simulations

This section analyzes the finite sample performance of the extremum estimator based on the second-order partial derivatives of the LCF from the previous section (labeled PD) and compares it to a third-order cumulant estimator (C3), a fourth-order cumulant estimator (C4), and the ordinary least squares estimator (OLS) using data generated from Monte-Carlo simulations.

Consider the errors-in-variables model,

$$X_1 = \alpha_1 + X_1^* + U_1 \tag{5a}$$

$$X_2 = \alpha_2 + X_2^* + U_2 \tag{5b}$$

$$Y = \alpha_Y + \beta_1 X_1^* + \beta_2 X_2^* + \varepsilon \tag{5c}$$

¹For more details see Ben-Moshe (2013).

Let $\beta_1 = 1$, $\beta_2 = 1$, and $(\alpha_1, \alpha_2, \alpha_Y) = (1, 1, 1)$.

Table 1 displays the means and standard errors (in parentheses) of the Monte Carlo distributions of the estimates of β_1 (the estimates of β_2 are similar) obtained from 100 simulations of sample size $N = 1,000$ without measurement error ($U_1 = U_2 = 0$) and with measurement error (U_1 and U_2 iid standard normal). In the first two columns (Design 1), (X_{1n}^*, X_{2n}^*) are generated by drawing $2N$ iid samples from a beta distribution with parameters $(1, 2)$ and adjusting the regressors to each have variance 2 and covariance 1 (by multiplying the generated variables by a Cholesky decomposition of the covariance matrix). In the third and fourth columns (Design 2), (X_{1n}^*, X_{2n}^*) are generated in the same way as the first two columns except that they are drawn from a χ^2 -distribution with 5 degrees of freedom. Similarly, in the fifth and sixth columns (Design 3), (X_{1n}^*, X_{2n}^*) are generated from a t-distribution with 5 degrees of freedom.

In all three designs, the OLS estimator has the tightest confidence bands around β_1 when there is no measurement error and is badly biased when there is measurement error. With measurement error, the PD estimator has the tightest confidence bands around β_1 . The t-distribution is symmetric so that there is no identifying information from the skewness, which manifests itself in biased C3 estimates with the largest standard errors.

Table I. Estimates of β_1

Estimator	Design 1: $\beta(1, 2)$		Design 2: $\chi^2(5)$		Design 3: $t(5)$	
	$U_k = 0$	$U_k \sim N(0, 1)$	$U_k = 0$	$U_k \sim N(0, 1)$	$U_k = 0$	$U_k \sim N(0, 1)$
PD	1.01 (0.07)	1.02 (0.16)	1.02 (0.06)	1.00 (0.10)	1.00 (0.13)	1.01 (0.12)
C3	1.02 (0.07)	1.02 (0.20)	1.01 (0.06)	1.01 (0.11)	1.05 (0.78)	0.83 (1.47)
C4	0.98 (0.13)	1.04 (0.30)	0.97 (0.15)	0.98 (0.21)	1.04 (0.16)	0.98 (0.19)
OLS	1.00 (0.03)	0.82 (0.04)	1.00 (0.03)	0.78 (0.03)	1.00 (0.03)	0.75 (0.03)

Estimates are based on the errors-in-variables model in Equations (5a)-(5c). For each design, we generate 100 simulations with 1,000 observations. The unobserved regressors are drawn from a $\beta(1, 2)$ distribution (Design 1), $\chi^2(5)$ distribution (Design 2) or $t(5)$ distribution (Design 3) with $\text{Var}(X_1^*) = \text{Var}(X_2^*) = 2$ and $\text{Cov}(X_1^*, X_2^*) = 1$. The first, third, and fifth columns have no measurement errors while the second, fourth, and sixth columns have measurement errors drawn from a standard normal distribution. The numbers without parentheses are means and the numbers in parentheses are standard errors.

Figures 1, 2, and 3 plot for Designs 1, 2, and 3 respectively the 5%, 50%, and 95% of the Monte-Carlo distributions of the estimates of the second-order partial derivatives,

$$\left. \frac{\partial^2 \varphi_{\mathbf{X},Y}(\mathbf{s})}{\partial s_1 \partial s_2} \right|_{(u,u,0)} \quad \left. \frac{\partial^2 \varphi_{\mathbf{X},Y}(\mathbf{s})}{\partial s_1 \partial s_3} \right|_{(u,u,0)} \quad \left. \frac{\partial^2 \varphi_{\mathbf{X},Y}(\mathbf{s})}{\partial s_2 \partial s_3} \right|_{(u,u,0)}$$

No constant function lies between the 5% and 95% confidence bands in all but one of the graphs in figures 1, 2, and 3 so we can be confident of identifying variation in all but one of the PDs (the exception is $\left. \frac{\partial^2 \varphi_{\mathbf{X},Y}(\mathbf{s})}{\partial s_1 \partial s_2} \right|_{(u,u,0)}$ when the regressors are generated from the t -distribution).

The t -distribution approaches the normal distribution as the degrees of freedom approach infinity. Figure 4 plots the second-order partial derivatives of a t -distribution with 10 degrees of freedom and although the median has variation, a constant function does lie between the 5% and 95% confidence bands so we are less confident about identification and our estimator.

Figures 1, 2, and 3 show a tradeoff between values of u close to the origin and values of u far from the origin; when u is close to the origin then the second-order partial derivatives are more accurately estimated (narrow confidence bands) but there is less variation in the second-order partial derivatives from their value at the origin while when u is far from the origin the second-order partial derivatives are less accurately estimated (wide confidence bands) but there is more variation in the second-order partial derivatives from their value at the origin.

We performed other simulations using the model in Equations (5a)-(5c) using various distributions, variances, and covariances, and choices of β (and using different ways to construct the dependence between X_1^* and X_2^* and different distributions of errors). In most the simulations, the PD estimator had the tightest confidence bands around β_1 . Further, we plotted the second-order partial derivatives of $\varphi_{\mathbf{X},Y}(\mathbf{s})$ to check for variation and the appropriateness of using the PD estimator.

4 Conclusion

This paper considers identification and estimation of coefficients in a multivariate linear regression with measurement errors in all the variables. Assuming that the unobserved regressors are not jointly normal and independent of errors, we identify and estimate the coefficients using variation in the second-order partial derivatives of the LCF of the unobserved regressors, which is only possible if the unobserved regressors are not jointly normal.

In our simulations, we find that the set on which $w(u) > 0$ is important but not the choice of function $w(u)$. This is similar to choosing bandwidth and could be an interesting topic for future research.

5 Appendix

The following relationship will be used in the proofs below.

$$\begin{aligned}
 \varphi_{\mathbf{X},Y}(s_1, \dots, s_K, s_y) &= \ln E \left[\exp \left(i \sum_{k=1}^K X_k s_k + i Y s_y \right) \right] \\
 &= \ln E \left[\exp \left(i \sum_{k=1}^K (X_k^* + U_k) s_k + i \left(\sum_{k=1}^K \beta_k X_k^* + \varepsilon \right) s_y \right) \right] \\
 &= \ln E \left[\exp \left(i \sum_{k=1}^K (s_k + \beta_k s_y) X_k^* + i \sum_{k=1}^K s_k U_k + i s_y \varepsilon \right) \right] \tag{6}
 \end{aligned}$$

where the second equality follows by substituting in $Y = \beta_1 X_1^* + \dots + \beta_K X_K^* + \varepsilon$ and $X_k = X_k^* + U_k$.

5.1 Proof of Lemma 2.1

Proof. Assume $(\mathbf{X}^*, \varepsilon)$ is dependent. Let $\tilde{\beta}_k = c_k \beta_k$ where $c_k \neq \{0, 1\}$, $\varphi_{\tilde{U}_k}(\cdot) \equiv 0$, and $\varphi_{\tilde{\mathbf{X}}^*, \tilde{\varepsilon}}(\cdot) = \varphi_{X_1^* + U_1, \dots, X_K^* + U_K, \varepsilon + \sum_{k=1}^K \beta_k (X_k^* (1 - c_k) - c_k U_k)}(\cdot)$. Then using (6),

$$\varphi_{\mathbf{X},Y}(\mathbf{s}) = \ln E \left[\exp \left(i \sum_{k=1}^K (s_k + \beta_k s_y) X_k^* + i \sum_{k=1}^K s_k U_k + i s_y \varepsilon \right) \right]$$

$$\begin{aligned}
&= \ln E \left[\exp \left(i \sum_{k=1}^K (s_k + c_k \beta_k s_y) (X_k^* + U_k) + i s_y \left(\varepsilon + \sum_{k=1}^K \beta_k (X_k^* (1 - c_k) - c_k U_k) \right) \right) \right] \\
&= \ln E \left[\exp \left(i \sum_{k=1}^K (s_k + \tilde{\beta}_k s_y) \tilde{X}_k^* + i s_y \tilde{\varepsilon} \right) \right] \\
&= \varphi_{\tilde{\mathbf{X}}^*, \tilde{\varepsilon}} \left(s_1 + \tilde{\beta}_1 s_y, \dots, s_K + \tilde{\beta}_K s_y, s_y \right) + \sum_{k=1}^K \varphi_{\tilde{U}_k} (s_k)
\end{aligned}$$

Hence, $\tilde{\beta}$ is observationally equivalent to β .

Assume (\mathbf{X}^*, U) is dependent. Let $\tilde{\beta}_k = c_k \beta_k$ where $c_k \neq \{0, 1\}$, $\varphi_{\tilde{\varepsilon}}(\cdot) = \varphi_{\varepsilon}(\cdot)$, and $\varphi_{\tilde{\mathbf{X}}^*, \tilde{U}}(\cdot) = \varphi_{c_1 X_1^*, \dots, c_K X_K^*, U_1 + X_1^* (1 - c_1), \dots, U_K + X_K^* (1 - c_K)}(\cdot)$. Then using (6),

$$\begin{aligned}
\varphi_{\mathbf{X}, Y}(\mathbf{s}) &= \ln E \left[\exp \left(i \sum_{k=1}^K (s_k + \beta_k s_y) X_k^* + i \sum_{k=1}^K s_k U_k + i s_y \varepsilon \right) \right] \\
&= \ln E \left[\exp \left(i \sum_{k=1}^K \left(s_k + \frac{\beta_k}{c_k} s_y \right) c_k X_k^* + i \sum_{k=1}^K s_k (U_k + X_k^* (1 - c_k)) + i s_y \varepsilon \right) \right] \\
&= \ln E \left[\exp \left(i \sum_{k=1}^K (s_k + \tilde{\beta}_k s_y) \tilde{X}_k^* + i \sum_{k=1}^K s_k \tilde{U}_k + i s_y \tilde{\varepsilon} \right) \right] \\
&= \varphi_{\tilde{\mathbf{X}}^*, \tilde{U}} \left(s_1 + \tilde{\beta}_1 s_y, \dots, s_K + \tilde{\beta}_K s_y, s_1, \dots, s_K \right) + \varphi_{\tilde{\varepsilon}}(s_y)
\end{aligned}$$

Hence, $\tilde{\beta}$ is observationally equivalent to β . □

5.2 Proof of Equivalences in Assumption 2.2

Proof. (i) \Leftrightarrow (ii). We prove the contrapositive. The unobserved vector of regressors \mathbf{X}^* is jointly normal with covariance matrix Σ if and only if

$$\varphi_{\mathbf{X}^*}(\boldsymbol{\omega}) = i \boldsymbol{\omega}' \boldsymbol{\mu} - \frac{1}{2} \boldsymbol{\omega}' \Sigma \boldsymbol{\omega} \quad (7)$$

By analytic continuity (e.g. Lukacs, 1970), the LCF of \mathbf{X}^* is given by Equation (7) if and only if the equation holds for some neighborhood of the origin. If (7) holds on a neighborhood around the origin then the second-order partial derivatives on this neighborhood are

$$\frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} = -\text{Cov}(X_{k_1}^*, X_{k_2}^*) \quad (8)$$

where $\text{Cov}(X_{k_1}^*, X_{k_2}^*) = - \left. \frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} \right|_{(\mathbf{0})}$ are the elements of Σ . By the fundamental theorem of calculus if (8) holds then (7) holds on this neighborhood (with possibly different mean $\boldsymbol{\mu}$) and so \mathbf{X}^* is normal.

(ii) \Leftrightarrow (iii). Using Equation (6) and Assumption 2.1,

$$\varphi_{\mathbf{X}, Y}(\mathbf{s}) = \varphi_{\mathbf{X}^*}(s_1 + \beta_1 s_y, \dots, s_K + \beta_K s_y) + \varphi_\varepsilon(s_y) + \sum_{k=1}^K \varphi_{U_k}(s_k)$$

The second-order partial derivatives are

$$\left. \frac{\partial^2 \varphi_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_{k_2}} \right|_{(\omega_1, \dots, \omega_K, 0)} = \frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} \quad 1 \leq k_1 < k_2 \leq K \quad (9)$$

$$\left. \frac{\partial^2 \varphi_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_y} \right|_{(\omega_1, \dots, \omega_K, 0)} = \sum_{k_2=1}^K \beta_{k_2} \frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} \quad 1 \leq k_1 \leq K \quad (10)$$

By (9), $\frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} \neq -\text{Cov}(X_{k_1}^*, X_{k_2}^*)$ if and only if $\left. \frac{\partial^2 \varphi_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_{k_2}} \right|_{(\omega_1, \dots, \omega_K, 0)} \neq c_{k_1 k_2}$ where $k_1 \neq k_2$. If $\frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} = -\text{Cov}(X_{k_1}^*, X_{k_2}^*)$ (and $\left. \frac{\partial^2 \varphi_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_{k_2}} \right|_{(\omega_1, \dots, \omega_K, 0)} = c_{k_1 k_2}$) then by (10), $\frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1}^2} \neq -\text{Var}(X_{k_1}^*)$ if and only if $\left. \frac{\partial^2 \varphi_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_y} \right|_{(\omega_1, \dots, \omega_K, 0)} \neq c_{k_1 k_2}$. \square

5.3 Proof of Theorem 2.1

Proof. Using Equation (6) and Assumption 2.1,

$$\varphi_{\mathbf{X}, Y}(\mathbf{s}) = \varphi_{\mathbf{X}^*}(s_1 + \beta_1 s_y, \dots, s_K + \beta_K s_y) + \varphi_\varepsilon(s_y) + \sum_{k=1}^K \varphi_{U_k}(s_k)$$

The second-order PDs are

$$\frac{\partial^2 \varphi_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_{k_2}} = \left. \frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} \right|_{(s_1 + \beta_1 s_y, \dots, s_K + \beta_K s_y)} \quad 1 \leq k_1 < k_2 \leq K \quad (11)$$

$$\frac{\partial^2 \varphi_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_y} = \sum_{k_2=1}^K \beta_{k_2} \left. \frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} \right|_{(s_1 + \beta_1 s_y, \dots, s_K + \beta_K s_y)} \quad 1 \leq k_1 \leq K \quad (12)$$

Define $\mathbf{b} := (b_1, \dots, b_K)$ and define

$$R_0^{k_1 k_2}(u; \mathbf{b}) := \left| \text{Cov}(X_{k_1}, X_{k_2}) + \left. \frac{\partial^2 \varphi_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_{k_2}} \right|_{(b_1 u, \dots, b_K u, -u)} \right|^2$$

$$= \left| \text{Cov}(X_{k_1}^*, X_{k_2}^*) + \frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} \Big|_{(u(b_1 - \beta_1), \dots, u(b_K - \beta_K))} \right|^2 \quad (13)$$

$$R_0^{k_1 y}(u; \mathbf{b}) := \left| \text{Cov}(X_{k_1}, Y) + \frac{\partial^2 \varphi_{\mathbf{X}, Y}(\mathbf{s})}{\partial s_{k_1} \partial s_y} \Big|_{(b_1 u, \dots, b_K u, -u)} \right|^2$$

$$= \left| \sum_{k_2=1}^K \beta_{k_2} \text{Cov}(X_{k_1}^*, X_{k_2}^*) + \sum_{k_2=1}^K \beta_{k_2} \frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} \Big|_{(u(b_1 - \beta_1), \dots, u(b_K - \beta_K))} \right|^2 \quad (14)$$

$$R_0(u; \mathbf{b}) := \sum_{1 \leq k_1 < k_2 \leq K} w_{k_1 k_2} R_0^{k_1 k_2}(u; \mathbf{b}) + \sum_{k_1=1}^K w_{k_1 y} R_0^{k_1 y}(u; \mathbf{b})$$

where (13) and (14) follow by substituting in (11) and (12). If $\mathbf{b} = \boldsymbol{\beta}$ then $R_0(u; \boldsymbol{\beta}) = 0$ for all $u \in \mathbb{R}$.

We now show that $\mathbf{b} = \boldsymbol{\beta}$ is the only solution to $R_0(u; \mathbf{b}) = 0$. If $R_0(u; \mathbf{b}) = 0$ then

$$\frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1} \partial \omega_{k_2}} \Big|_{(u(b_1 - \beta_1), \dots, u(b_K - \beta_K))} = -\text{Cov}(X_{k_1}^*, X_{k_2}^*)$$

$$\frac{\partial^2 \varphi_{\mathbf{X}^*}(\boldsymbol{\omega})}{\partial \omega_{k_1}^2} \Big|_{(u(b_1 - \beta_1), \dots, u(b_K - \beta_K))} = -\text{Var}(X_{k_1}^*)$$

where the first equality follows by Equation (13) and the second equality follows by using the first equality and Equation (14). Using Assumption 2.2, the only solution is $\mathbf{b} = \boldsymbol{\beta}$. Hence, $\boldsymbol{\beta} = \underset{\mathbf{b} \in \mathcal{B}}{\text{argmin}} \int_{\mathbb{R}} R_0(u; \mathbf{b}) w(u) du$. \square

References

- BEN-MOSHE, D. (2013): “Identification and Estimation of Factor Models Using the Log Characteristic Function,” working paper.
- BEN-MOSHE, D. (2014): “Identification of Dependent Nonparametric Distributions in a System of Linear Equations,” working paper.
- CARROLL, R.J., RUPPERT, D., STEFANSKI, L.A., CRAINICEANU, C. (2006): *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition* (Chapman and Hall).
- CHEN, X., HU, Y. and LEWBEL, A. (2009): “Nonparametric Identification and Estimation of Nonclassical Errors-in-Variables Models Without Additional Information”, *Statistica Sinica*, 19, 949-968.
- CHEN, X., HONG, H. and NEKIPELOV, D. (2011): “Nonlinear Models of Measurement Errors,” *Journal of Economic Literature*, 49(4), 901-937.
- CHEN, X., HONG, H. and TAMER, E. (2005): “Measurement Error Models with Auxiliary Data,” *Review of Economic Studies*, 72(2), 343-366.
- CHENG, C.L. and VAN NESS, J. W. (1999): *Statistical Regression with Measurement Error* Kendall’s Library of Statistics 6. Arnold, London.

- CRAGG, J.G. (1997): "Using higher moments to estimate the simple errors-in-variables model," *The RAND Journal of Economics*, 28, S71-S91.
- DAGENAIS, M. G. and DAGENAIS, D.L. (1997): "Higher moment estimators for linear regression models with errors in the variables," *Journal of Econometrics*, 76, 193-221.
- DELAIGLE, A., HALL, P. and MEISTER, A. (2008): "On Deconvolution with Repeated Measurements," *Annals of Statistics*, 36(2), 665-685.
- ERICKSON, T., JIANG, C., and WHITED, T. M. (2013): "Minimum Distance Estimation of the Errors-in-Variables Model Using Linear Cumulant Equations," *Journal of Econometrics*, forthcoming.
- ERICKSON, T. and WHITED, T. M. (2002): "Two-step GMM Estimation of the Errors-in-variables Model Using High-order Moments," *Econometric Theory*, 18, 776-799.
- ERICKSON, T. and WHITED, T. M. (2012): "Treating Measurement Error in Tobin's q ," *Review of Financial Studies*, 25, 1286-1329.
- FULLER, W. (1987): "Measurement Error Models," Wiley, New York.
- GEARY, R.C. (1942): "Inherent relations between random variables," *Proceedings of the Royal Irish Academy*, 47, 63-76.
- GEARY, R.C. (1949): "Sampling aspects of the problem from the error-in-variable approach," *Econometrica*, 17, 26-28.
- GILLARD, J.W. (2010): "An overview of linear structural models in errors in variables regression," *Revstat*, 8(1), 57-80.
- HAUSMAN, J.A., NEWEY W.K., and POWELL J.L., (1995): "Nonlinear errors in variables: estimation of some Engel curves," *Journal of Econometrics*, 65, 205-233.
- HSIAO, C. and WANG, Q.K., (2000): "Estimation of structural nonlinear errors-in-variables models by simulated least-squares method," *International Economic Review*, 41(2), 523-542.
- HU, Y. (2008): "Identification and Estimation of Nonlinear Models with Misclassification Error using Instrumental Variables: A General Solution," *Journal of Econometrics*, 144, 27-61.
- HU, Y. and RIDDER, G. (2012): "Estimation of Nonlinear Models with Mismeasured Regressors Using Marginal Information," *Journal of Applied Econometrics*, 27, 347-385.
- HU, Y. and SCHENNACH, S. M. (2007): "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 76, 195-216.
- KHATRI, C. G. and RAO, C. R. (1968): "Solutions to Some Functional Equations and their Applications to Characterization of Probability Distributions," *Sankhyā*, 30, 167-180.
- KHATRI, C. G. and RAO, C. R. (1972): "Functional Equations and Characterization of Probability Laws Through Linear Functions of Random Variables," *Journal of Multivariate Analysis*, 2, 162-173.
- KLEPPER, S. and LEAMER, E. (1984): "Consistent sets of estimates for regressions with errors in all variables," *Econometrica*, 52, 163-183.
- KOTLARSKI, I. (1967): "On Characterizing the Gamma and Normal Distribution," *Pacific Journal of Mathematics*, 20, 69-76.
- LEWBEL, A. (1997): "Constructing Instruments for Regressions With Measurement Error When no Additional Data are Available, with An Application to Patents and R&D,"

- Econometrica*, 65(5), 1201-1213.
- LEWBEL, A. (2012): "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models," *Journal of Business & Economic Statistics*, 30, 67-80.
- LI, T. (2002): "Robust and Consistent Estimation of Nonlinear Errors-in-Variables Models," *Journal of Econometrics*, 110, 1-26.
- LUKACS, E. (1970): "Characteristic functions (2nd Edition)," New York: Hafner Pub. Co.
- NEWHEY, N. and MCFADDEN, D. (1994): "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 36, 2111-2245.
- PAL, M. (1980): "Consistent moment estimators of regression coefficients in the presence of errors in variables," *Journal of Econometrics*, 14, 349-364.
- POLLARD, D. (1984): *Convergence of Stochastic Processes* (New York: Springer).
- RAO, B. L. S. P. (1992): "Identifiability in Stochastic Models". Academic Press, Boston.
- RAO, C. R. (1971), "Characterization of Probability Laws by Linear Functions," *Sankhyā*, 33, 265-270.
- REIERSØL, O. (1950): "Identifiability of a linear relation between variables which are subject to error," *Econometrica* 9, 1-24.
- SCHENNACH, S. M. (2004a): "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 72(1), 33-75.
- SCHENNACH, S. M. (2004b): "Nonparametric Estimation in the Presence of Measurement Error," *Econometric Theory* 20, 1046-1093.
- SCHENNACH, S. M. and HU, Y. (2013): "Nonparametric Identification and Semiparametric Estimation of Classical Measurement Error Models Without Side Information," *Journal of the American Statistical Association*, 108, 177-186.
- VAN MONTFORT, K., MOOIJAART, A., DE LEEUW, J. (1987): "Regression with errors in variables: estimators based on third order moments," *Statistica Neerlandica* 41, 223-238.
- VAN MONTFORT, K., MOOIJAART, A., DE LEEUW, J. (1989): "Estimation of regression coefficients with the help of characteristic functions," *Journal of Econometrics* 41, 267-278.
- WANSBEEK, T. J., MEIJER, E. (2000): *Measurement error and latent variables in econometrics*, Elsevier, Amsterdam.
- WILLASSEN, Y. (1979): "Extension of Some Results by Reiersol to Multivariate Models," *Scandinavian Journal of Statistics* 6, 89-91.

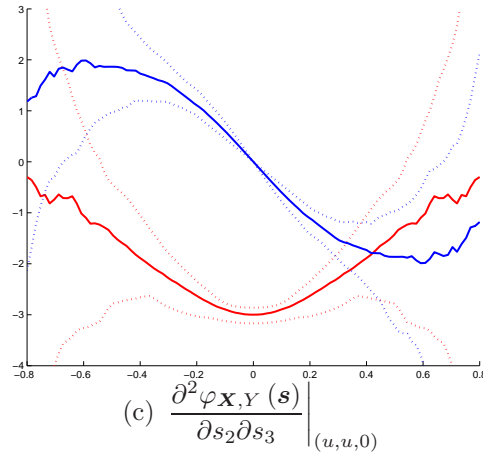
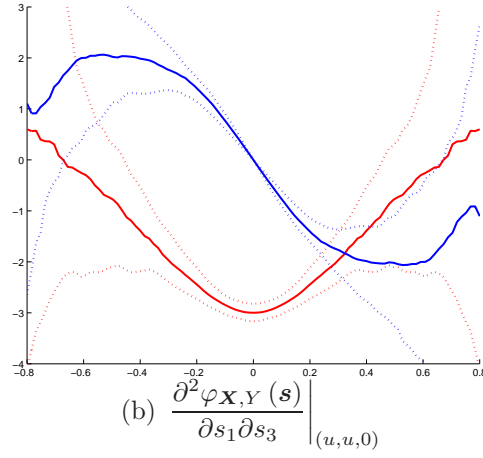
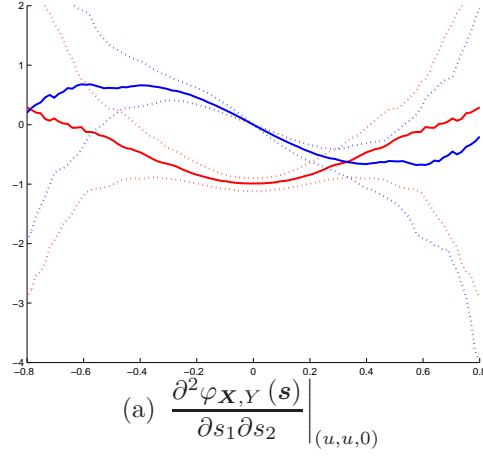


Figure 1: Estimates from the errors-in-variables model with (X_1^*, X_2^*) drawn from a χ^2 -distribution with 5 degrees of freedom and adjusted so that the regressors each have variance 2 and covariance 1. The graphs plot the second-order partial derivatives of $\varphi_{\mathbf{X},Y}$ where the red lines are the real parts and the blue lines are the imaginary parts. The solid lines are the medians of the Monte-Carlo draws. The dotted lines are the 5% - 95% confidence bands. No constant function can fit between the confidence bands suggesting variation in all the second-order partial derivatives.

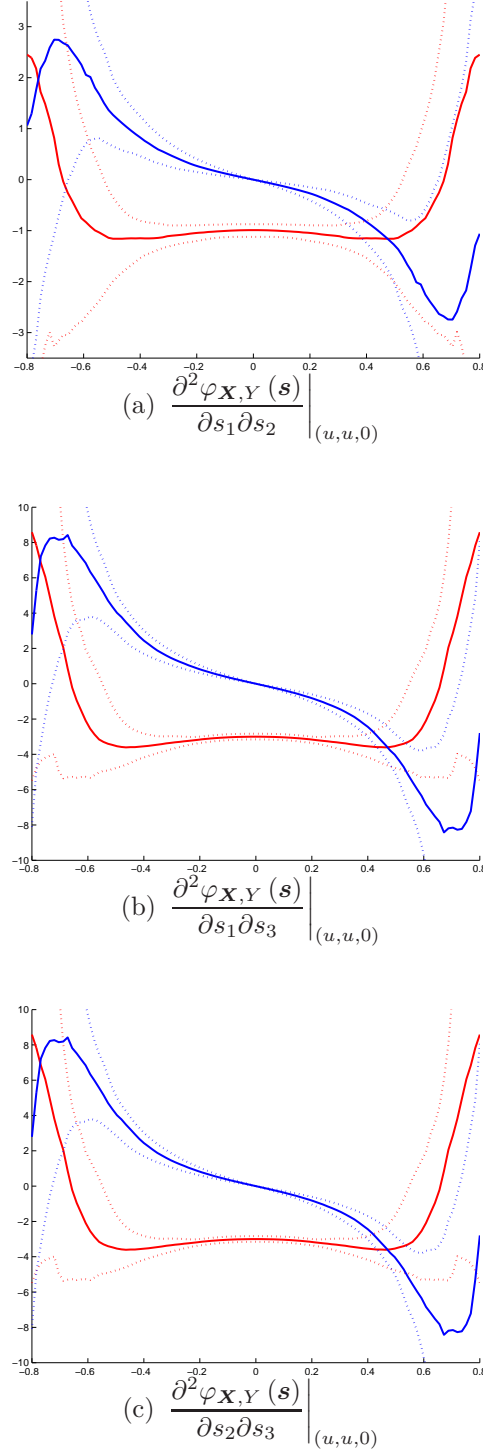


Figure 2: Estimates from the errors-in-variables model with (X_1^*, X_2^*) drawn from a β -distribution with parameters $(\alpha, \beta) = (1, 2)$ and adjusted so that the regressors each have variance 2 and covariance 1. The graphs plot the second-order partial derivatives of $\varphi_{\mathbf{X},Y}$ where the red lines are the real parts and the blue lines are the imaginary parts. The solid lines are the medians of the Monte-Carlo draws. The dotted lines are the 5% - 95% confidence bands. No constant function can fit between the confidence bands in graphs (b) and (c) suggesting variation in these second-order partial derivatives.

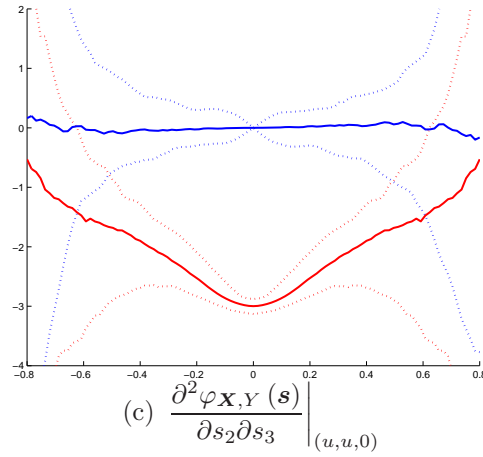
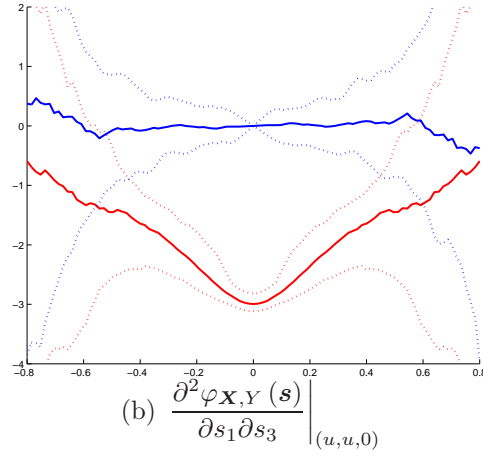
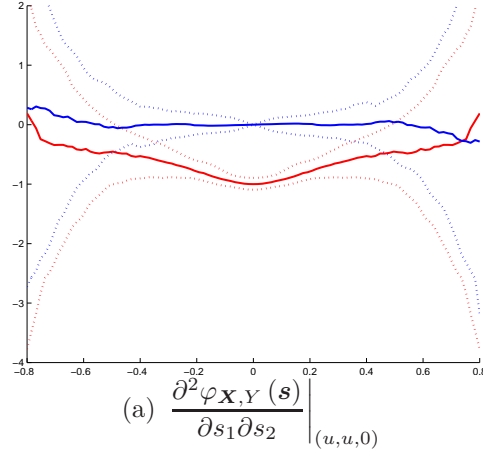


Figure 3: Estimates from the errors-in-variables model with (X_1^*, X_2^*) drawn from a **t-distribution with 5 degrees of freedom** and adjusted so that the regressors each have variance 2 and covariance 1. The graphs plot the second-order partial derivatives of $\varphi_{\mathbf{X},Y}$ where the red lines are the real parts and the blue lines are the imaginary parts. The solid lines are the medians of the Monte-Carlo draws. The dotted lines are the 5% - 95% confidence bands. No constant function can fit between the confidence bands suggesting variation in all the second-order partial derivatives.

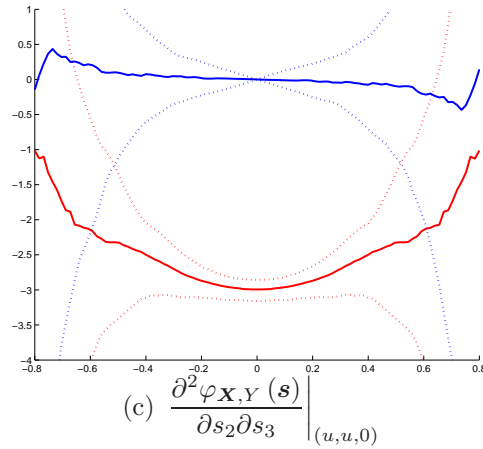
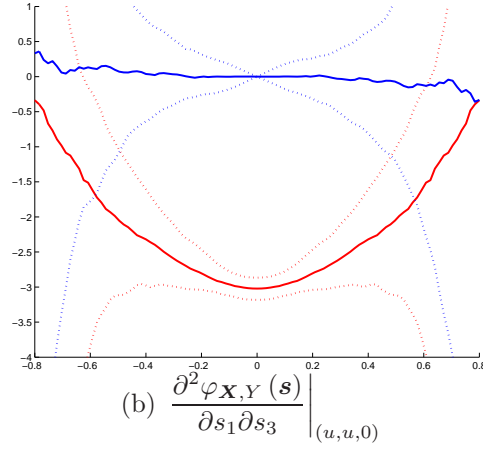
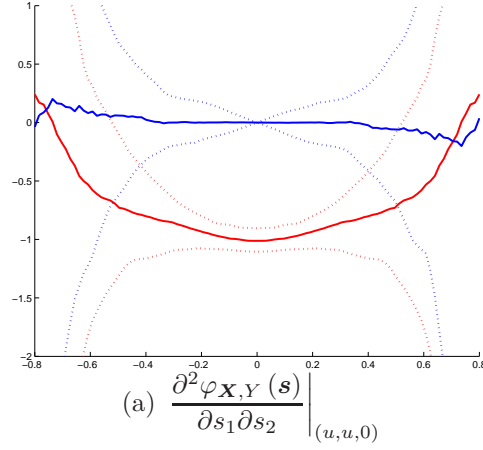


Figure 4: Estimates from the errors-in-variables model with (X_1^*, X_2^*) drawn from a **t-distribution with 10 degrees of freedom** and adjusted so that the regressors each have variance 2 and covariance 1. The graphs plot the second-order partial derivatives of $\varphi_{\mathbf{X},Y}$ where the red lines are the real parts and the blue lines are the imaginary parts. The solid lines are the medians of the Monte-Carlo draws. The dotted lines are the 5% - 95% confidence bands. A constant function can fit between the confidence bands suggesting that there may not be variation in all the second-order partial derivatives.